NIST's Performance Evaluation System

Paul Doremus
Director, NIST Program Office

Visiting Committee on Advanced Technology

September 13, 2004

NIST's Stakeholder Community

- External stakeholders that request information about NIST's performance:
  - Administration
    - Department of Commerce (DOC)
    - Office of Management and Budget (OMB)
    - Office of Science and Technology Policy (OSTP)
  - Congress
    - House and Senate Appropriations Committees
    - House and Senate Authorizing Committees
  - Federal Advisory Committees
    - NIST-wide: Visiting Committee on Advanced Technology (VCAT)
    - External Boards for each of NIST's extramural programs

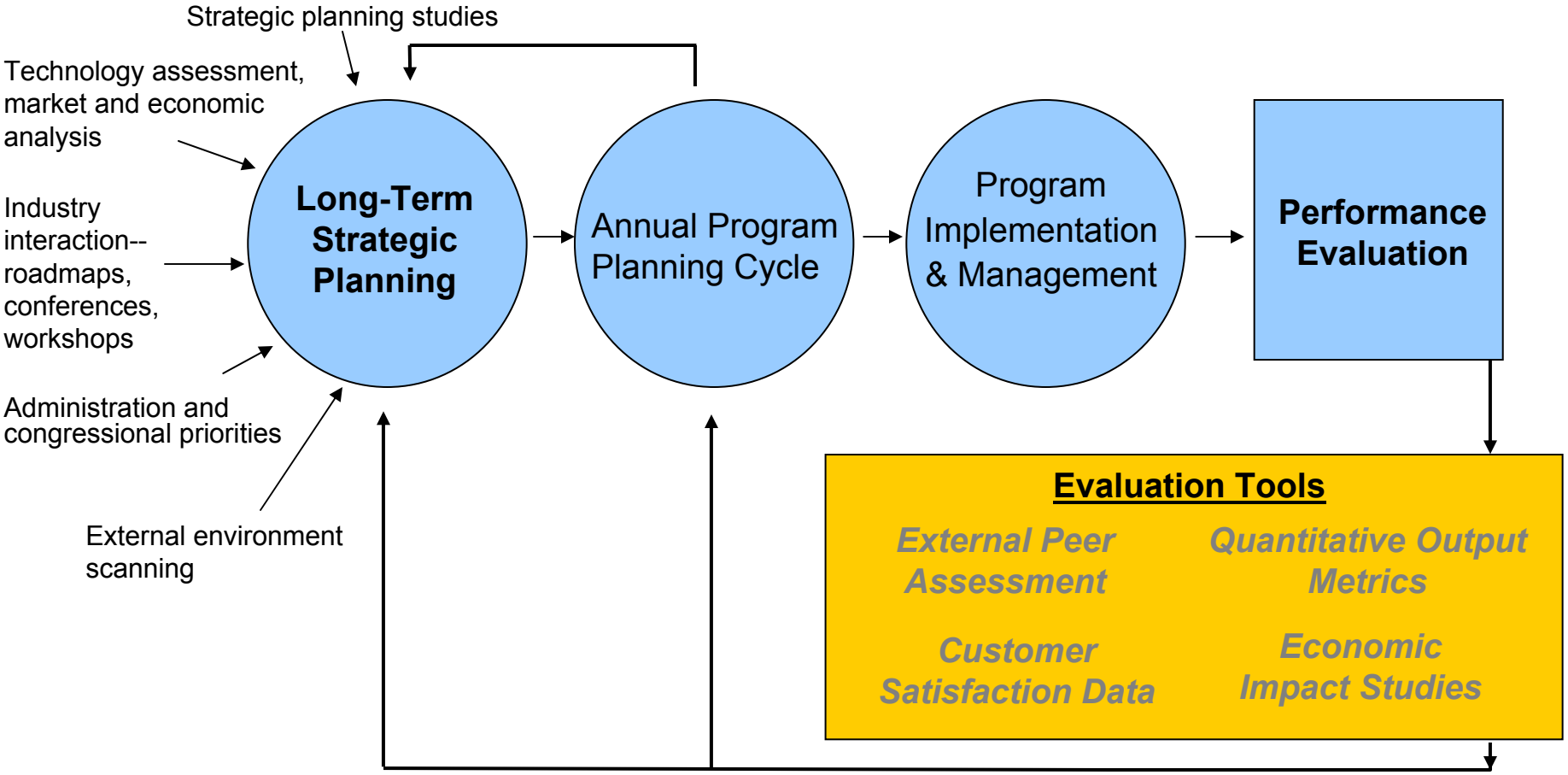
External Requirements for Performance Evaluation

- Stakeholder requirements for performance-based management
  - Legislative: Government Performance and Results Act (GPRA, 1993)
    - Requires all agencies to produce strategic plans every three years and annual performance plans and reports
    - Primary focus: "Vital few, outcome-oriented measures"
  - Executive: President's Management Agenda
    - Budget and performance integration
    - R&D investment criteria
    - Program Assessment Rating Tool
    - Primary focus: good management practices, proven results

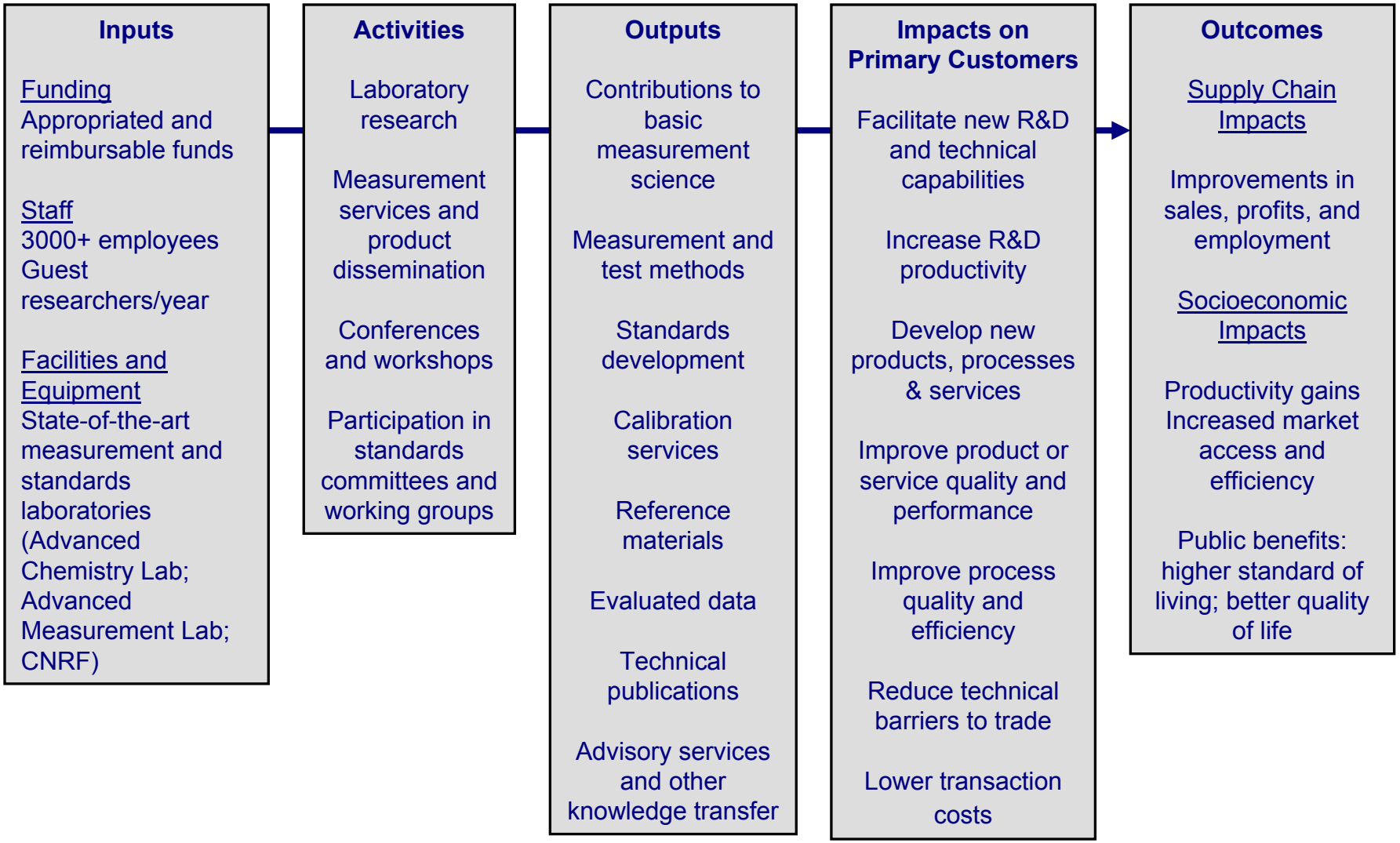
Issues in Reporting Results

- Mission requires complex evaluation system
- No "vital few outcome-oriented metrics"
- Performance data not synchronized with budget cycle
- Uncertain utility for stakeholders

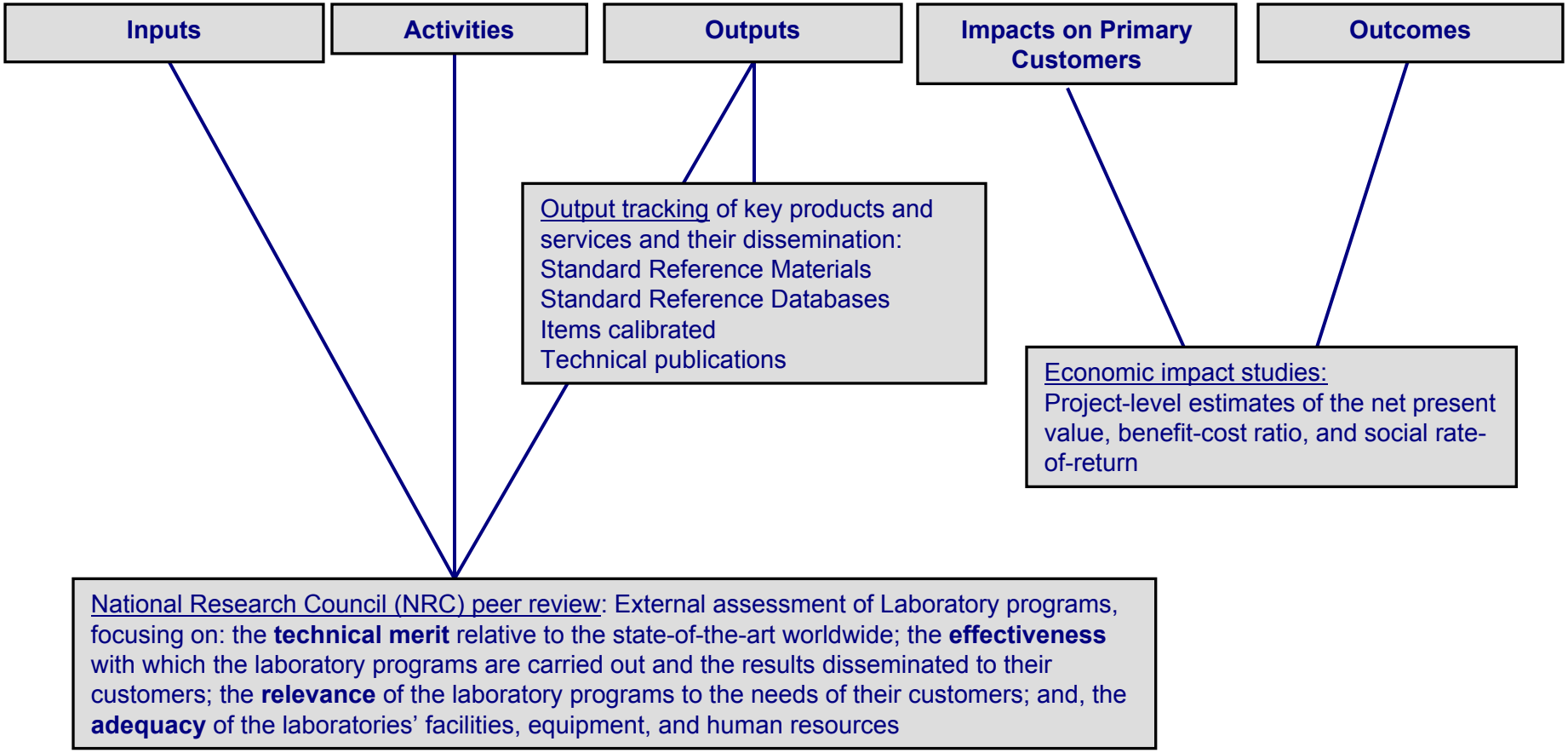
NIST's Planning and Performance Evaluation System



NIST's Value Chain



NIST Laboratories Impact Path and Evaluation Tools



Strengths & Weaknesses of Measurement Methods

	Scope & Purpose	Strengths	Limitations
Peer Review	Assess technical quality within operating units. Provides essential data for quality control, laboratory management & planning.	Broad and detailed review by external technical experts. Balanced panels; expertise matches each operating unit. NRC independence, high technical capability, and internal quality controls.	Intrinsic features of peer review: panel judgments are not quantifiable; observations and findings are highly contextual and detailed; assessments are not comparable (e.g. no cumulative performance ranking).
Quanti-tative Output Metrics	Diverse output indicators for key functions. Important to track for internal management & resource planning.	Direct counts of activities and outputs generate highly reliable quantitative data. Robust data collection systems. Data are cumulative and allow trend analysis.	Provide no information on quality or impact; trends require contextual interpretation; indicators not uniformly relevant to all OUs; indicators as a set are not a comprehensive output meaasure.
Impact Studies of Research Outcomes	Assess down-stream impacts of research projects & infratechnologies. Provides data for evaluating research outcomes & long-term planning.	Provides quantitative and qualitative data re. outcomes. Provides data on impacts over long time periods and across layers of the supply chain affected by NIST. Highly qualified economists and technical specialists conduct detailed analyses using well-developed research methods.	Studies are intermittent and results are not cumulative; elements of user population often are too diffuse to measure; uneven availability and quality of industry data; methodological problems specific to each measure; outcomes are specific to each project (limited comparability); studies are expensive.

Internal Needs for Performance Evaluation

- Evaluate current work portfolio
  - Technical merit relative to state-of-the-art
  - Relevance to customer needs
- Evaluate retrospective performance
  - Test validity of impact logic model
  - Improved understanding of impact pathways

Performance-Based Management

- Scorecard management system developed in FY 2004, being refined for FY 2005
- Scorecard framework:
  - **Programs**: Maximize impact and demonstrate programmatic results
  - **People**: Continuous improvement in developing and managing a high-performance workforce
  - **Resources**: Maximize operational, organizational and programmatic leverage and impact
  - **Customers**: Improve customer satisfaction with NIST's products and services

Issues in Evaluating Research Outcomes

- Scope of measurable impact
  - Impact pathways within and across supply chains
  - Causal complexity and attribution challenges
    - Retrospective: how to assess counterfactual results?
    - Prospective: how to value alternative investments?
  - Boundaries of quantitative and qualitative assessment
- Difficulty of assessing research portfolios
  - Portfolio composition: mix and complementarity
  - Portfolio balance: risk and time horizons
- Data access, quality, and reliability
- Long time frames

Issues in Evaluating Research Investments

- Intrinsic measurement challenges
  - Evaluating productivity of laboratory functions
  - Validating scientific and technical judgment
- Qualitative evaluation most useful internally, but not entirely persuasive externally
- Consequent approach:
  - Diverse set of measurement methods
  - Most comprehensive method: qualitative peer assessment provided by the National Research Council